

Researcher Centred Data Repository Workflows

Potential Workflows



Abstract

Supporting the management of research data in universities requires a range of services and components, including a long-term vehicle for archiving and sharing data. Such data repositories can serve multiple functions, and ensuring they engage with researchers through appropriate workflows is a key challenge. Data.uel, a new data repository at the University of East London, seeks to provide workflows for four use cases appropriate to institutions with a smaller research base.

It provides researcher-centred workflows for depositing datasets, which can happen in advance of an overarching description of a research project or following it. Accordingly, the repository has a data model that permits the creation of a record for a single or multiple assets, and for records to be related to siblings as well as to parents. The data may also be deposited before or after a related publication is added to the publications repository (ROAR).

A separate data repository was created for two reasons: to customise ingest to relate closely to scholarly workflows, and to enable the addition of data that was not Open. We know that not all data can or should be Open, so allow researchers to make their data available in a choice of three ways: Open Access, using open licensing; Available on Request, where the data owner would approve release case-by-case; and Closed to all but depositor and administrator, where data has to be kept for compliance/ evidentiary reasons but is not suitable for sharing.

Technical

Both ROAR and data.uel are hosted and maintained by the University of London Computer Centre (ULCC). UEL created the specification and designed the look and feel, while ULCC built the back end repository.

ULCC created the linking system to allow for the workflows discussed above. This involved two different types of linking between ePrint records, and one external link. As the two types of ePrint to ePrint link do not exist in stock EPrints, some customisation was needed.

There are 3 types of links deployed:

- 1. Project to Datasets**
This was accomplished using the Containers plugin for EPrints. The plugin allows a record to be created which can contain any number of dataset records in the repository.

The linking is the first part of the Project workflow. The depositor can search the repository by dataset title, linking as many as needed.

These links are then stored in the Container ePrint.
- 2. ePrints to ePrints - data.uel to ROAR**
This is a custom plugin written by ULCC. It allows Datasets to link to the Publications which they support (held in ROAR). The link is initiated in the data.uel workflow. The depositor searches for a title in ROAR, which autofills live from the ROAR repository.

This is a many to many relationship, so the depositor can link to multiple publications in ROAR.

The link is only stored in data.uel, so ROAR has to look up data.uel to get the links. ROAR sends a GET request for a JSON, to see if a link has been made.

If data.uel tells ROAR that there has been no link, it gets an empty file.

If a link(s) is present, the details are returned in the JSON.

The link is displayed on the summary page, with the colour of the destination repository designating where it is going to. This information is also in the machine readable metadata as:

```
<meta name="eprints.internal_link_link" content="http://roar.uel.ac.uk/id/eprint/3550" />
```

- 3. Data to external datasets (registry)**
Links to external resources are handled by the existing EPrints field linked_resource, with minor customisation.

It allows for externally stored datasets (UK Data Archive, Figshare, etc.) to link to an ePrint. This means it is possible to have a metadata-only record (registry) which links to the actual data.

The depositor inputs the URL, the title and a description of what the object is. These are then displayed in the same way as the repository links above, but with a different colour, to help differentiate them.

Authors:

Stephen Grace
Research Services Librarian
s.grace@uel.ac.uk

David McElroy
RDM Officer
d.mcelroy@uel.ac.uk

Research Data Services
University of East London

Acknowledgements:

We would like to thank Rory McNicholl and Timothy Miles-Board of University of London Computer Centre for the excellent work developing and supporting our repositories.

